View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Illinois Digital Environment for Access to Learning...

A STUDY OF DURATION IN SPEECH PRODUCTION

Marilyn Bereiter

Temporal organization in speech production is a function of the brain's transmission of nervous impulses to the articulatory mechanisms. At this time, it is only a matter of conjecture how or in what sequence the brain "realizes" an idea and then processes lexical, syntactic and phonological requirements before executing a single utterance. It is with the execution of an utterance or a string of utterances that we are concerned here.

An utterance is a sequence of phonological units associated with a time dimension. The smallest phonological unit is a single phoneme which can be considered the most basic speech component. The question being studied is which phonological unit (phoneme, syllable, word, etc.) constitutes the minimal unit of brain transmission. If the phoneme (each with a unique duration) was the basic unit, an impulse would be sent from the brain which would activate articulation of that phoneme. Another impulse would then be transmitted in the same manner, and another -- until the entire sequence of phonemes had been uttered. This would assume that the duration of an utterance would be the sum of the durations of the phonemic components. Studies have shown that if a phoneme is slightly altered in duration, an adjacent phoneme will compensate by adjusting its duration in order to ensure an apparently "scheduled" total duration (Kozhevnikov and Chistovich, 1965).

The problem now is to establish the level or phonological unit in which temporal compensation takes place (Lehiste, 1970). This would determine the minimal unit of speech production. Two of the methods of investigation have yielded two non-compatible conclusions, but each contain statistical validity. Lehiste (1970) showed that steady (/stɛd+i/) actually had a shorter duration than stead (/stɛd/) even though there was an additional phoneme/syllable. In one experiment she was able to conclude that neither the phoneme nor the syllable could be the minimal unit of speech production, since in both cases a longer duration for steady would have been anticipated. In this experiment, the domain of temporal compensation seemed to be the word level.

Kozhevnikov and Chistovich (1965) examined the relationships among

the different phonological units with respect to carefully controlled variation in the rate of speech. They assumed the concept of an articulatory program which directs the production of speech segments without temporal consideration. Then the rate of speech characterizes the speed at which the articulatory program is realized. The results of their experiment indicated that with a change in tempo, ratios among phonemes vary significantly (in quick tempo, there was even evidence of total vowel reduction); however, syllables and words retained their same ratios regardless of the rate of speech. Their conclusion was that the articulatory program is based on the syllable as the minimal unit of production.

The work that follows is based on the work of Kozhevnikov and Chistovich (1965) following the premise that the syllable is the minimal unit of speech production. The study involves repetition of a single sentence at different tempos and a statistical analysis of the ratios between each phonological level.

The sample sentence was: The tiger pounced onto the streaking chimpanzee. This sentence breaks down into 32 phonemes, 12 syllables, 7 words and 2 phrases (see Figure 1). Care was taken in choosing a sentence that contained distinct and unambiguous segmental boundaries. Lehiste (1972) discussed the perceptual reality of segmentation and concluded that the production and perception of timing patterns is relative to changes in manner of articulation. These changes are manifested quite clearly in the visual display of the acoustic waveform. With this in mind, it was desirable to intentionally alternate voiced and voiceless segments. The sample sentence also had to be long enough so as to minimize the effects of intonation. Since I wanted to determine relationships among units at different levels, the sentence had to contain more than one phrase as well as polysyllabic words and polyphonemic syllables.

There was only one subject who spoke the sentence three times each at three different tempos, beginning with a normal rate of speech, then quickening the pace and finally decreasing the tempo to a sub-normal rate. The tempo intervals were rather arbitrary and subjective; a discussion of their repercussions will be included later. The subject spoke directly into a microphone that was connected to a Siemens Oscillomink. The signal was amplified and then broken down into components and analyzed. The

analysis was then simultaneously charted on calibrated paper at the rate of 10 cm/sec. In order to determine the duration of each segment, the acoustic waveforms as well as the amplitude displays had to be studied.

By measuring the length of each segment represented on the oscillograms, I was able to quantify the data in units of seconds. I had nine separate samples belonging to three (hopefully) distinct groups or populations. Since I was interested in the ratios between phonological units of the same structural level, I calculated a percent for each unit representing that unit's fraction of the next higher level. For example, in Table 1a, P_1 represents 34.6% of S_1 in the slow group while P_2 represents 65.4% of S_1. In Table 1c, W_1 is 16.0% of Ph_1 and W_2 is 84.0% of Ph_1 also in the slow group. A mean was first computed from the raw data (in seconds) and then the percent was determined rather than first computing the percents and then taking a mean percent. This was done in order to minimize variance within the groups in an effort to consider the groups as representing three distinct speeds rather than a continuum.

After the percents had been calculated, I was interested in analyzing how well correlated each level (phonological unit) of each tempo was with the other tempos of that level. A multivariant analysis gave the following correlation coefficients:

UNIT	FAST:NORMAL	NORMAL:SLOW	FAST:SLOW
phoneme	0.977	0.960	0.975
syllable	0.999	0.977	0.996
word	0.999	0.998	0.999
phrase	1.000	1.000	1.000

The results indicated that each group was so highly correlated that something must have been askew. Though the correlations do improve (approach +1) as the phonological units increase (which would be anticipated by my hypothesis), it was impossible to draw any valid conclusions from this test. I decided to go back a step to inveatigate why all of the groups were so highly correlated.

By examining the standard deviations of the means (in seconds) for each tempo (Table 2), the overlap between groups indicates that they are

not at all distinct; and in fact, most groups (especially at the phoneme level) belong to the same population. Therefore, it is not at all surprising to find the high correlation coefficients, since one would expect a coefficient of +1 when correlating a group with itself.

I would like to say that we could look for a few tendencies or generalities relating to the original hypothesis; but given the questionable nature of these statistics, anything deduced from them would lack validity. However, I cannot refrain from pointing out one interesting observation: in Table 2 as the level increases, there exists a seemingly significant increase in the distinctness of the groups. This is perhaps due to the very essence of the problem. Each group at the sentence level is unquestionably (statistically) and empirically distinct. With each reduction in level (decreasing the block of time that the component units comprise), the amount of variability also is reduced. At the phoneme level there is only one possible variable that could contribute to the tempo discrimination. According to Kozhevnikov and Chistovich (1965), certain phonemes (especially consonants) have very discrete duration limits whereas the vowels can increase or decrease in length almost infinitely. I had hoped to be able to point to a particular level and say that it was the minimal unit of production.

The inconclusiveness of this study is not due to a wrong hypothesis but rather to a non-rigorous experimental procedure. (The hypothesis may or may not be correct but it is impossible to make any judgment based on these data.) Instead, a more precise timing mechanism must be implemented to ensure the distinctness of each tempo. Kozhevnikov and Chistovich (1965) used a buzzer that sounded when the utterance was to have been ended. This enabled the subjects to pace themselves; and with some practice, a normalized tempo for each rate of speech could be achieved. With such a mechanism, the sample size could easily be increased which would further ensure a more valid statistical analysis.

Depending on the results of a future experiment testing the same hypothesis, further investigation should be focused towards solving the discrepancy between the results Lehiste (1970) achieved with her method of experimentation and those deduced by Kozhevnikov and Chistovich (1965). One possible procedure could include Lehiste's general approach of com-

paring pairs of words like stead and steady but incorporating them into a longer utterance. This would minimize any distortions in duration that might be caused by intonation of words in isolation.

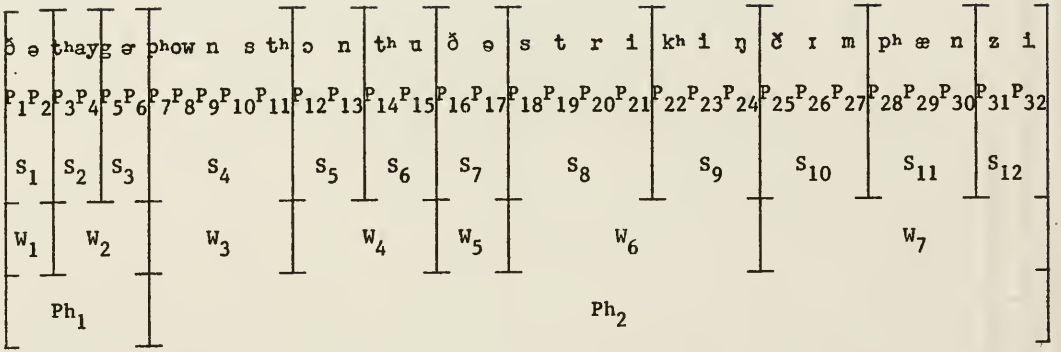

Figure 1.

Segmentation of the sentence: The tiger pounced onto the streaking chimpanzee.

Key: P = phoneme
S = syllable
W = word
Ph = phrase

Table 1a.

Phoneme Percent of Syllable for Three Rates of Speech

	FAST	NORMAL	SLOW
P_1	36.1	33.0	34.6
P_2	63.9	67.0	65.4
P_3	56.0	51.1	57.8
P_4	44.0	48.9	42.2
P_5	44.1	40.5	40.4
P_6	55.9	59.5	59.6
P_7	27.7	24.4	27.5
P_8	30.4	34.3	33.0
P_9	11.3	11.7	09.3
P_{10}	18.5	15.7	18.7
P_{11}	12.1	13.9	11.4
P_{12}	68.8	69.5	65.4
P_{13}	31.2	30.5	34.6
P_{14}	62.4	60.8	66.5
P_{15}	37.6	39.2	33.5
P_{16}	41.7	46.2	50.9
P_{17}	58.3	53.8	49.1
P_{18}	49.3	44.6	49.1
P_{19}	13.7	15.3	12.4
P_{20}	11.1	14.0	09.5
P_{21}	25.9	26.1	28.9
P_{22}	46.4	40.1	43.0
P_{23}	36.7	36.3	42.0
P_{24}	16.9	23.6	15.0
P_{25}	51.4	60.2	49.5
P_{26}	20.2	23.4	19.1
P_{27}	28.3	16.4	31.3
P_{28}	28.4	27.6	25.2
P_{29}	45.7	43.1	43.2
P_{30}	25.8	29.3	31.6
P_{31}	36.9	36.7	34.8
P_{32}	63.1	63.3	65.2

Table 1b.

Syllable Percent of Word for Three Rates of Speech

	FAST	NORMAL	SLOW
S_1	100.0	100.0	100.0
S_2	65.6	64.0	60.3
S_3	34.4	36.0	39.7
S_4	100.0	100.0	100.0
S_5	54.6	56.3	54.5
S_6	45.4	43.7	45.5
S_7	100.0	100.0	100.0
S_8	53.3	51.9	51.0
S_9	46.7	48.1	49.0
S_{10}	29.0	26.7	28.8
S_{11}	45.5	47.4	45.2
S_{12}	25.5	25.9	25.9

Table 1c.

Word Percent of Phrase for Three Rates of Speech

	FAST	NORMAL	SLOW
W_1	16.6	17.2	16.0
W_2	83.4	82.8	84.0
W_3	21.0	22.4	23.6
W_4	13.1	13.7	13.9
W_5	05.4	05.1	05.1
W_6	22.6	23.1	21.4
W_7	38.0	35.7	36.0

Table 1d.

Phrase Percent of Sentence for Three Rates of Speech

	FAST	NORMAL	SLOW
Ph_1	18.2	18.5	17.4
Ph_2	81.8	81.5	82.6

Table 2a.

Mean Duration and Standard Deviation of Phonemes for Three Rates of Speech

	no overlap*	FAST $\bar{X}$**	FAST sd	NORMAL $\bar{X}$	NORMAL sd	SLOW $\bar{X}$	SLOW sd
P_1		.030	.010	.033	.012	.037	.006
P_2		.053	.006	.067	.031	.070	.010
P_3		.153	.015	.157	.015	.197	.006
P_4		.120	.017	.150	.000	.143	.012
P_5		.063	.012	.070	.010	.090	.010
P_6	x	.080	.000	.103	.012	.133	.006
P_7		.130	.010	.140	.010	.207	.006
P_8	x	.143	.006	.197	.006	.247	.006
P_9		.053	.006	.067	.021	.070	.010
P_{10}		.087	.032	.090	.010	.140	.000
P_{11}		.057	.021	.080	.026	.087	.006
P_{12}		.110	.010	.137	.015	.157	.006
P_{13}		.050	.010	.060	.010	.083	.012
P_{14}		.083	.015	.093	.015	.133	.015
P_{15}		.050	.017	.060	.010	.067	.015
P_{16}		.050	.010	.060	.010	.083	.012
P_{17}		.070	.000	.070	.000	.080	.010
P_{18}		.133	.021	.137	.032	.170	.010
P_{19}		.037	.006	.047	.015	.043	.006
P_{20}		.030	.000	.043	.015	.033	.006
P_{21}		.070	.000	.080	.017	.100	.010
P_{22}		.110	.010	.113	.012	.143	.006
P_{23}		.087	.015	.103	.015	.140	.010
P_{24}		.040	.010	.067	.015	.050	.010
P_{25}		.133	.012	.147	.023	.163	.006
P_{26}		.050	.000	.057	.006	.063	.006
P_{27}		.070	.000	.090	.010	.103	.006
P_{28}		.110	.017	.120	.010	.130	.017
P_{29}		.143	.065	.187	.015	.223	.015
P_{30}		.100	.010	.127	.021	.163	.031
P_{31}		.080	.010	.089	.006	.103	.006
P_{32}		.137	.012	.150	.020	.193	.021

* indicates each group is a separate population

** in seconds

Table 2b.

Mean Duration and Standard Deviation of Syllables for Three Rates of Speech

	no overlap	FAST $\bar{X}$	FAST sd	NORMAL $\bar{X}$	NORMAL sd	SLOW $\bar{X}$	SLOW sd
S_1		.083	.015	.100	.040	.107	.015
S_2		.273	.025	.307	.015	.340	.010
S_3	x	.143	.012	.173	.015	.223	.012
S_4	x	.470	.010	.573	.049	.750	.000
S_5	x	.160	.010	.197	.012	.240	.010
S_6	x	.133	.012	.153	.006	.200	.020
S_7		.120	.010	.130	.010	.163	.006
S_8		.270	.020	.307	.025	.347	.015
S_9	x	.237	.015	.283	.012	.333	.025
S_{10}		.253	.012	.293	.029	.330	.010
S_{11}	x	.387	.015	.433	.015	.517	.051
S_{12}		.217	.015	.237	.025	.297	.023

Table 2c.

Mean Duration and Standard Deviation of Words for Three Rates of Speech

	no overlap	FAST $\bar{X}$	FAST sd	NORMAL $\bar{X}$	NORMAL sd	SLOW $\bar{X}$	SLOW sd
W_1		.083	.015	.100	.040	.107	.015
W_2	x	.417	.015	.480	.020	.563	.006
W_3	x	.470	.010	.573	.049	.750	.000
W_4	x	.293	.015	.350	.010	.440	.026
W_5		.120	.010	.130	.010	.163	.006
W_6	x	.507	.035	.590	.026	.680	.035
W_7	x	.857	.031	.963	.064	1.143	.042

Table 2d.

Mean Duration and Standard Deviation of Phrases for Three Rates of Speech

	no overlap	FAST $\bar{X}$	FAST sd	NORMAL $\bar{X}$	NORMAL sd	SLOW $\bar{X}$	SLOW sd
Ph_1	x	.500	.010	.580	.035	.670	.020
Ph_2	x	2.247	.080	2.607	.125	3.177	.049

Table 2e.

Mean Duration and Standard Deviation of Sentence for Three Rates of Speech

	no overlap	FAST $\bar{X}$	FAST sd	NORMAL $\bar{X}$	NORMAL sd	SLOW $\bar{X}$	SLOW sd
Sentence	x	2.747	.075	3.187	.129	3.847	.051

References

Kozhevnikov, V.A. and L.A. Chistovich. 1965. Speech: Articulation and Perception. Translated by J.P.R.S., Washington, D.C., No. JPRS 30,543. Moscow-Leningrad. pp. 69-118.

Lehiste, Ilse. 1970. "Temporal Organization of Spoken Language". Working Papers in Linguistics, No. 4. Ohio State University. pp 96-114.

Lehiste, Ilse. 1972. "Manner of Articulation, Parallel Processing, and the Perception of Duration". Working Papers in Linguistics, No. 12. Ohio State University. pp. 33-52.